Abstract

Estimation of DSGE model with or without filter

Lee, Young-hwan
Department of Economics
The Graduate School
Seoul National University

So far, various filtering techniques have been introduced to the field of economic studies since it has been conventional method to use pre-filtered data to estimate the DSGE model which is constructed to explain business cycle fluctuation. To name a few of widely used filter, Hodrick-Prescott filter, band-pass filter, and Beveridge-Nelson filter. Such diversity of methodological alternatives is a possible source of confusion if the detailed mechanisms of filtering methods and the differences between them are remained unanswered and ambiguous. Furthermore, as Fabio Canova(1998) points out, the results of estimation are not independent of filtering method. Therefore, in this study, the relationship between filter and estimation will be studied. Specifically, it will focus on i) clarify and codify the tacit assumptions on commonly used filtering method, ii) identify the effect of filtering on estimation, iii) propose the estimation method without filtering.

Keywords : Filtering, DSGE model estimation, Frequency domain analysis

Student Number : 2010-23026

Contents

1 Introduction

When we want to estimate the parameterized DSGE model which is constructed to explain business cycle fluctuation of an economy, it is troublesome to find the relevant observation since econometricians cannot observe the business cycle component directly. So far, to tackle this problem, filtering was widely used technique. Hodrick-Prescott filter, high-pass filter, and band-pass filter are typical examples. However, as Fabio Canova(1998) points out, the results of estimation are not robust but vary by filtering methods. It is the fundamental motivation of this study.
Actually, it is natural to have a doubt that pre-filtered data based estimation is distortive since, by definition, filtering is a set of actions those are taken to artificially suppress and amplify certain characteristics of a given data set. More precisely, if estimation is based on the pre-filtered data, it is not possible to identify the information that is supposed to be contained in the suppressed component. At the same time, estimation will exaggerate the effect of amplified characteristics. Therefore, it is trivial that filtering can cause distortion in estimation.
Let's get back to the DSGE model estimation issue. The problem is the fact that business cycle component is not separately observable but only the totality of business cycle component and growth component is available source of information[1]. Therefore, we can extract the business cycle component from observation if a filter can perfectly suppress the trend and growth component. It is the ideal case that is not attainable in most practices. Now, the issue is to find the characteristics that we can distinguish business cycle from observation. Hodrick-Prescott filter, high-pass filter, and many other filters are implicitly assuming that the power of business cycle component is concentrated on certain area in frequency domain and therefore we can discern business cycle component from growth component. At a glance this assumption looks valid. However, it is problematic if we are considering estimation. Observe that it is equivalent to assume the prior knowledge on business cycle component as well as growth component and recall that the business cycle is the component of interest. Suppose that we are quite sure that our prior knowledge on business cycle component is correct. Then es-

[1]It is similar problem to the colored noise problem. For the more interesting topics on colored noise, refer Jesús et al. (2009).

timation itself is meaningless process because it is more reasonable to stick to that belief. Suppose not. Then filtering is based on possibly wrong assumptions so that estimation is not properly work. i.e. we are presupposing arbitrary values on the parameters when we use a filter. Therefore, filtering itself can severely distort the result of estimation. To solve this problem, this paper has threefold objective. i) clarify and codify the tacit assumptions of widely used filtering method, ii) identify the effect of filtering on estimation, iii) propose the estimation method without filtering. This paper is organized as follows. In section 2, several mathematical formulation will be represented. Section 3 contains simple analysis on filtering problem. In section 4, numerical example will be studied. Section 5 concludes the study.

2 Preliminary

To specify the data generating process, consider the data generating process as below. Put

$$y_t^{ob} = y_t^{gr} + y_t^{cy}$$

Where

$$\begin{aligned} y_t^{ob} &= \text{Observed variable} \\ y_t^{gr} &= \text{Growth component} \\ y_t^{cy} &= \text{Business cycle component} \end{aligned}$$

All of them are stochastic processes those are defined on $t \in [1, T]$ and its dimension is $l \times 1$. From now on, we will stick to the notations those are defined in this section.

Suppose that we have a log linearized DSGE model which models the business cycle fluctuation of an economy. Let just assume that it is parameterized by $\theta \in \Theta$. It means that we have a pdf for $\{y_t^{cy}\}_{t=1}^T$ that is a set of observation of business cycle components from time 1 to T. i.e. we have

$$L(\theta|\{y_t^{cy}\}_{t=1}^T)$$

Such function can be derived or approximated by using Kalman filter in time domain or CLT in frequency domain. However, econometricians are troubled since $\{y_t^{cy}\}_{t=1}^T$ is not observable. Actually, the totality of growth,

and business cycle component is the only available source of information. Therefore, it is not feasible estimation that use $L(\theta|\{y_t^{cy}\}_{t=1}^T)$ directly. However, we have some clue if we can assume that we already have some prior knowledge or belief on business cycle component and growth component. Assuming the existence such prior belief or knowledge is innocuous. It is supported by the following argument.
Suppose that there is no recognizable difference that we can discern $\{y_t^{cy}\}_{t=1}^T$ from $\{y_t^{ob}\}_{t=1}^T$. Then, it is pointless to claim that we should not use observation instead of business cycle component. On the other hand, suppose that we have no idea what is the difference between the observation and the business cycle component. In this case, we have no choice but to use observation itself. Consequently, any endeavors to discern business cycle component from observation assumes certain prior knowledge or belief on the nature of business cycle component and observation[2].
Based on the available observation and prior knowledge on it, we can handle this problem in two different ways.

i) Use $\{\tilde{y}_t^{cy}\}_{t=1}^T$ instead of $\{y_t^{cy}\}_{t=1}^T$ where $\{\tilde{y}_t^{cy}\}_{t=1}^T$ is a sort of estimator of business cycle component given observation. For example, use

$$L(\theta|E[\{y_t^{cy}\}_{t=1}^T|\{y_t^{ob}\}_{t=1}^T])$$

ii) Based on the prior knowledge on stochastic nature of data generating process, derive $P(\{y_t^{cy}\}_{t=1}^T|\{y_t^{ob}\}_{t=1}^T)$ and consider $\{y_t^{cy}\}_{t=1}^T$ as latent variable. i.e use

$$E\left[L(\theta|\{y_t^{cy}\}_{t=1}^T)|\{y_t^{ob}\}_{t=1}^T\right]$$

Actually, i) is equivalent to using pre-filtered data. and ii) is the method what I want to propose in this paper which does not use any filtering on data. In next part, the prior knowledge on data generating structure is formalized and parameterized.

[2]Leaving aside a discussion of the rightness or wrongness of this belief, it is important to codify the tacit assumptions implied by such endeavors. If discussions proceed from the ambiguous premises, it could be meaningless to talk about truthness of argument. Recall that clarifying such ambiguity is one of the purpose of this study.

2.1 Setting

For the growth component, put

$$y_t^{gr} = y_t^{gr_0} + y_t^{gr_1}$$

Where

$$\begin{aligned} y_t^{gr_0} &= \text{Deterministic growth component} \\ y_t^{gr_1} &= \text{Stochastic growth component} \end{aligned}$$

The first part of growth component describes the deterministic growth that is known to or believed by econometricians to be true with certainty. Different from first part, second part describes the growth part that is known to econometricians but lacks certainty except that mean is zero for all t. i.e. prior knowledge of econometrician is too limited to pin down exact value of it. Therefore what we can do at best is to pin down the stochastic structure. For example, what can be said is whether it is of low frequency or smooth function of time t. As an illustrative example, consider the case belows where $l = 1$.

$$y_t^{gr} = \alpha y_{t-1}^{gr} + \mu + \varepsilon_t$$

Where ε is white noise, $|\alpha| < 1$, $y_0^{gr} = 0$. Then

$$y_t^{gr} = \frac{1-\alpha^t}{1-\alpha}\mu + \sum_{k=0}^{t-1} \alpha^k \varepsilon_{t-k}$$

and

$$y_t^{gr_0} = \frac{1-\alpha^t}{1-\alpha}\mu \qquad , \qquad y_t^{gr_1} = \sum_{k=0}^{t-1} \alpha^k \varepsilon_{t-k}$$

Or, if we assume linear trend and stochastic part be a colored noise v^c, then

$$y_t^{gr_0} = \mu t \qquad , \qquad y_t^{gr_1} = v_t^c$$

where $|\mu| < \infty$.

To model the first part, it is sufficient to assume any function in $l^2[1, T]$ for our purpose. Let this part be fully parameterized by parameter μ. For the second part, the stochastic growth, it takes a little more effort to model the stochastic nature. In this study, we put

$$y_t^{gr_1} = \sum_{n=0}^{\frac{T}{2}-1} X_n \cos\left(\frac{2\pi n}{T}t\right) + Y_n \sin\left(\frac{2\pi n}{T}t\right)$$

where

$$\begin{pmatrix} X_n \\ Y_n \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_X(n|\mu) & \Sigma_{XY}(n|\mu) \\ \Sigma_{XY}(n|\mu) & \Sigma_Y(n|\mu) \end{pmatrix}\right)$$

It is quite general formulation that can cover many cases since

$$\text{span}\left\{\cos\left(\frac{2\pi n}{T}t\right), \sin\left(\frac{2\pi n}{T}t\right)\right\}_{n\in[0,\frac{T}{2}-1]} = l^2[1,T]$$

Note that y^{gr} is fully parameterized by μ if we assume that variance covariance matrices also are parameterized by μ. Put μ_0 be true parameter and is known to researchers.
For business cycle component, according to the Sims(2001) notation, put

$$y_t^{cy} = \sum_{j=0}^{k} A_j(\theta) S_{t-j}^{cy}$$

Where S_t^{cy} is $m \times 1$ vector of state variables which cause business cycle fluctuation and possibly not observed by econometricians. And $\{A_j(\theta)\}_{j=1}^k$ is a set of $m \times m$ matrices parameterized by θ. S_t^{cy} evolves as belows

$$S_{t+1}^{cy} = \Phi_1^{cy}(\theta) S_t^{cy} + \Phi_0^{cy}(\theta) \varepsilon_{t+1}^{cy}$$

where $\Phi_1^{cy}(\theta), \Phi_0^{cy}(\theta)$ parameterized by θ and ε^{cy} random shock that cause business cycle fluctuation. Let true θ is set to θ_0 In addition, assume that $\{X_n, Y_n\} \perp \{\varepsilon^{cy}\}$. i.e. the independence of y^{gr} and y^{cy} is assumed.

2.2 Data in frequency domain

In this paper, discussions are based on frequency domain since it makes the filtering problems more accessible. At this part, the frequency domain representation will be introduced. For convenience, let just assume that T is an even number. Then, $\{e^{i\frac{2\pi n}{T}t}\}_{n\in I}$ is a complete orthogonal basis of $l^2[1,T]$ where $I = \{n \in \mathbb{Z} | -\frac{T}{2} < n < \frac{T}{2}\}$ and $i = \sqrt{-1}$.

Write

$$\omega_n = \frac{2\pi n}{T}$$

And we define

$$\hat{y}(\omega_n) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} y_t e^{i\omega_n t}$$

for some $y \in l^2[1, T]$ and call $\{\hat{y}(\omega_n)\}_{n \in I}$ be the frequency domain representation of y. Note that $\hat{y}(\omega_n)$ is the coefficient of an orthogonal decomposition of y i.e.

$$y_t = \sum_{n \in I} \hat{y}(\omega_n) \sqrt{\frac{2\pi}{T}} e^{-i\omega_n t}$$

Observe that

i) the mapping $y \to \hat{y}$ is linear.

ii) $\hat{y}(\omega_{-n}) = \hat{y}(\omega_n)^*$ where * represent the complex conjugate.

From now on, frequency domain representation of each components will be derived. Since $y \to \hat{y}$ is linear,

$$\hat{y}^{ob}(\omega_n | \mu_0, \theta_0) = \hat{y}^{gr}(\omega_n | \mu_0) + \hat{y}^{cy}(\omega_n | \theta_0)$$

and

$$\begin{aligned} \hat{y}^{gr}(\omega_n | \mu_0) &= \hat{y}^{gr_0}(\omega_n | \mu_0) + \hat{y}^{gr_1}(\omega_n | \mu_0) \\ &= \hat{y}^{gr_0}(\omega_n | \mu_0) + \sqrt{\frac{T}{2\pi}} Z_n \end{aligned}$$

where

$$Z_n = \frac{1}{2} X_n + \frac{i}{2} Y_n$$

since we can easily check the below fact by using Euler's formula

$$y_t^{gr_1} = \sum_{n \in I} Z_n e^{i\omega_n t}$$

For the business cycle component,

$$\begin{aligned}\hat{y}^{cy}(\omega_n|\theta_0) &= \left(\sum_{j=0}^{k} A_j(\theta_0)e^{i\omega_n j}\right) S^{cy}(\omega_n|\theta_0) \\ &= \left(\sum_{j=0}^{k} A_j(\theta_0)e^{i\omega_n j}\right) \left(I - \Phi_1^{cy}(\theta_0)e^{i\omega_n}\right)^{-1} \Phi_0^{cy}(\theta_0)\hat{\varepsilon}^{cy}(\omega_n)\end{aligned}$$

Note that the frequency domain representation of each component also is random variable. Their elementary stochastic natures are as follows.

$$\{\hat{y}^{gr}(\omega_n|\mu_0)\}_{n\in I}\perp\{\hat{y}^{cy}(\omega_n|\theta_0)\}_{n\in I}$$

and

$$E\hat{y}^{gr}(\omega_n) = \hat{y}^{gr_0}(\omega_n), \quad E\hat{y}^{cy}(\omega_n) = 0$$

Define

$$\begin{aligned}\Sigma^{gr}(\omega_n|\mu_0) &= E\hat{y}^{gr_1}(\omega_n)\hat{y}^{gr_1}(\omega_n)^{\dagger} \\ \Sigma^{cy}(\omega_n|\theta_0) &= E\hat{y}^{cy}(\omega_n)\hat{y}^{cy}(\omega_n)^{\dagger} \\ \hat{y}^{ob_0}(\omega_n) &= \hat{y}^{ob}(\omega_n) - \hat{y}^{gr_0}(\omega_n) \\ \Sigma^{ob}(\omega_n|\mu_0,\theta_0) &= E\hat{y}^{ob_0}(\omega_n)\hat{y}^{ob_0}(\omega_n)^{\dagger} \\ &= \Sigma^{gr}(\omega_n|\mu_0) + \Sigma^{cy}(\omega_n|\theta_0)\end{aligned}$$

where † represents the conjugate transpose.
Furthermore, $\{\hat{y}^{gr_1}(\omega_n|\mu_0)\}_{n\in I}\{\hat{y}^{cy}(\omega_n|\theta_0)\}_{n\in I}$ is known to be complex gaussian asymptotically[3]. So that we can write

$$\begin{pmatrix}\hat{y}^{gr}(\omega_n|\mu_0) \\ \hat{y}^{cy}(\omega_n|\theta_0)\end{pmatrix} \sim CN\left(\begin{pmatrix}\hat{y}^{gr_0}(\omega_n) \\ 0\end{pmatrix}, \begin{pmatrix}\Sigma^{gr}(\omega_n|\mu_0) & 0 \\ 0 & \Sigma^{cy}(\omega_n|\theta_0)\end{pmatrix}\right)$$

And they are mutually independent either in either I^+ or I^- where $I^+ = \{n \in I | n \geq 0\}$ and $I^- = I - I^+$. From above fact, given observations and assumptions on growth component, we can derive the likelihood function on

[3] James Hannan (1970)

the parameter of interest θ as belows[4].

$$L(\theta|\{\hat{y}^{ob}(\omega_n)\}_{n\in I},\mu_0) = \prod_{n\in I^+} \frac{1}{\pi^l \det(\Sigma^{ob}(\omega_n|\mu_0,\theta)} \times e^{-tr[(\Sigma^{ob}(\omega_n|\mu_0,\theta))^{-1}\hat{y}^{ob_0}(\omega_n)\hat{y}^{ob_0}(\omega_n)^\dagger]}$$

2.3 Filter in frequency domain

In general, the purpose of filter is to take desirable component from the mixture of desirable and undesirable component. In this case, desirable component is y^{cy}. From now on, I will only consider the linear filter case to keep the analysis tractable. Actually, it does not seriously damage the generality of analysis when we confine the scope to linear filter since most filters what we use are fall into this category. Linear filter means a set of filters that acts like a linear operator. Suppose $\Lambda : l^2 \to l^2$ be a linear operator. We can represent a linear filter as a linear operator such that

$$\Lambda y^{ob} = \tilde{y}^{cy}$$

where $\tilde{y}^{cy}$ represents the filtered data.

Proposition 1. *A linear filter Λ can be represented by a set of $l \times l$ matrices $\{M_n\}_{n\in I}$*

Proof. Since $\{e^{i\omega_n t}\}_{n\in I}$ is a complete orthogonal basis in l^2, there exists a set of functions $\{M_n\}_{n\in I}$ such that $M_n : \mathbb{C}^l \to \mathbb{C}^l$ is equivalent to operator Λ. Observe that Λ is linear if and only if M_n is linear. i.e. M_n is $l \times l$ matrix. Therefore, we can find such $\{M_n\}$ that is equivalent to linear filter.■

Let $\tilde{y}^{cy}$ be a set of filtered data. Then we can say that

$$\tilde{y}^{cy}_t = \Lambda y^{ob}_t = \sum_{n\in I} M_n \hat{y}^{ob}(\omega_n)\sqrt{\frac{2\pi}{T}} e^{-i\omega_n t}$$

i.e. we are considering $M_n\hat{y}^{ob}(\omega_n)$ instead of $\hat{y}^{cy}(\omega_n)$ in frequency domain if we use filtered data for the estimation. Observe below proposition.

Proposition 2. *Suppose that we are trying to estimate the business cycle fluctuation parameter θ based on the filtered data. Suppose we are using a*

[4] Since $\hat{y}(\omega_{-n}) = \hat{y}(\omega_n)^*$

linear filter Λ which can be represented by $\{M_n\}_{n\in I}$. Then, we are considering likelihood function as below

$$
\begin{aligned}
L(\theta|\{\hat{y}^{ob}(\omega_n)\}_{n\in I}, \mu_0) &= \prod_{n\in I^+} \frac{1}{\pi^l \det(\Sigma^{cy}(\omega_n|\theta))} \\
&\times \quad e^{-tr[\Sigma^{cy}(\omega_n|\theta)^{-1} M_n^2 \hat{y}^{ob}(\omega_n)\hat{y}^{ob}(\omega_n)^\dagger]}
\end{aligned}
$$

Observe that this likelihood function is not equal to the one derived from previous part. This inconsistency is the underlying motive of this study. Based on the terminologies and findings from this section, we will delve into the effect of filter on estimation in next section.

3 Analysis

Recall that the filtering we want is a sort of point estimation on $\{\hat{y}^{cy}(\omega_n)\}_{n\in I}$. In this section, firstly, distortion on likelihood function that caused by filter is observed. And next, minimum mean squared error estimator of $\{\hat{y}^{cy}(\omega_n)\}_{n\in I}$ and its effect on estimation will be considered. After a simple derivation, we will realize that true $\{\hat{y}^{gr_0(\omega_n|\mu_0)}, \Sigma^{gr}(\omega_n|\mu_0), \Sigma^{cy}(\omega_n|\theta_0)\}_{n\in I}$ is assumed to be known otherwise minimum mean squared error estimator is not fixed. i.e. a filter is tacitly imposing certain assumptions on $\{\hat{y}^{gr_0}(\omega_n), \Sigma^{gr}(\omega_n|\mu_0), \Sigma^{cy}(\omega_n|\theta_0)\}_{n\in I}$ if we consider the filter as a mean squared error estimator. Based on this observation, tacit assumptions of widely used filters will be investigated in the second part.
Finally, the equivalence of filtering to EM algorithm is derived in the third part. It is noteworthy that using pre-filtered data is equivalent to apply EM algorithm with only one iteration if sufficient regularity conditions are hold.

3.1 Distortion caused by pre-filtered data

Suppose that we are using a filter which can be represented by $\{M_n\}_{n\in I}$. We have pointed out in previous section that pre-filtered data based estimation is equivalent to use below likelihood function if the prior belief on growth component is fixed at $\mu = \mu_0$.

$$L^{\text{filtered}}(\theta|\{\hat{y}^{ob}(\omega_n), M_n\}_{n\in I}, \mu_0) = \prod_{n\in I^+} \frac{1}{\pi^l \det(\Sigma^{cy}(\omega_n|\theta))} \times e^{-tr[\Sigma^{cy}(\omega_n|\theta)^{-1}M_n^2\hat{y}^{ob}(\omega_n)\hat{y}^{ob}(\omega_n)^\dagger]}$$

However, the true likelihood function is

$$L(\theta|\{\hat{y}^{cy}(\omega_n)\}_{n\in I}, \mu_0) = \prod_{n\in I^+} \frac{1}{\pi^l \det(\Sigma^{cy}(\omega_n|\theta))} \times e^{-tr[\Sigma^{cy}(\omega_n|\theta)^{-1}\hat{y}^{cy}(\omega_n)\hat{y}^{cy}(\omega_n)^\dagger]}$$

So that we can observe the distortion of likelihood function caused by filtering, we need to find the measurement of distortion. Write

$$D_n(\theta|M_n) = e^{-tr[\Sigma^{cy}(\omega_n|\theta)^{-1}(M_n^2\hat{y}^{ob}(\omega_n)\hat{y}^{ob}(\omega_n)^\dagger - \hat{y}^{cy}(\omega_n)\hat{y}^{cy}(\omega_n)^\dagger)}$$

and put

$$d_n(\theta|M_n) = \log D_n(\theta|M_n)$$

Then,

$$\log\left(\frac{L^{\text{filtered}}(\theta|\{\hat{y}^{ob}(\omega_n), M_n\}_{n\in I}, \mu_0)}{L(\theta|\{\hat{y}^{cy}(\omega_n)\}_{n\in I}, \mu_0)}\right) = \sum_{n\in I^+} d_n(\theta|M_n)$$

Its conditional expectation given $\{\hat{y}^{ob}(\omega_n)\}_{n\in I}$ is as belows

$$\begin{aligned}
E\left[\sum_{n\in I^+} d_n(\theta|M_n)|\hat{y}^{ob}(\omega_n)\right] &= \\
E&\left[\sum_{n\in I^+} -tr[\Sigma^{cy}(\omega_n|\theta)^{-1}(M_n^2\hat{y}^{ob}(\omega_n)\hat{y}^{ob}(\omega_n)^\dagger - \hat{y}^{cy}(\omega_n)\hat{y}^{cy}(\omega_n)^\dagger)]|\hat{y}^{ob}(\omega_n)\right] \\
&= -\sum_{n\in I^+} tr[\Sigma^{cy}(\omega_n|\theta)^{-1}(M_n^2\hat{y}^{ob}(\omega_n)\hat{y}^{ob}(\omega_n)^\dagger - E[\hat{y}^{cy}(\omega_n)\hat{y}^{cy}(\omega_n)^\dagger)|\hat{y}^{ob}(\omega_n)]] \\
&= -\sum_{n\in I^+} tr[\Sigma^{cy}(\omega_n|\theta)^{-1}(M_n^2\hat{y}^{ob}(\omega_n)\hat{y}^{ob}(\omega_n)^\dagger \\
&\qquad -(\Sigma^{cy}(\omega_n|\theta)^{-1}\Sigma^{ob}(\omega_n|\mu_0,\theta))^2\hat{y}^{ob_0}(\omega_n)\hat{y}^{ob_0}(\omega_n)^\dagger)]
\end{aligned}$$

From above equation, we can find that expected distortion on likelihood function caused by filter can exist. In reality, it exists in most cases. Below is direct application of this observation.

Proposition 3. *If and only if $\hat{y}^{ob}(\omega_n) = \hat{y}^{ob_0}(\omega_n)$ for all $n \in I$, we can find the filter $\{M_n\}_{n\in I}$ such that expected distortion is zero at a given point θ^* for all possible observation. Specifically, $M_n = \Sigma^{cy}(\omega_n|\theta^*)^{-1}\Sigma^{ob}(\omega_n|\mu_0, \theta^*)$, expected distortion on likelihood function is zero at $\theta = \theta^*$*

Actually, such $\{M_n\}_{n\in I}$ represent the minimum mean squared error estimator of y^{cy} given assumption that $\theta = \theta^*$. It is explored in detail at next part.

3.2 Minimum mean squared error filtering

Proposition 4. *$\{M_n\}_{n\in I}$ represent minimum mean squared estimator*[5] *if we define*

$$M_n = \Sigma^{cy}(\omega_n|\theta_0)\left(\Sigma^{ob}(\omega_n|\mu_0, \theta_0) + \hat{y}^{gr_0}(\omega_n)\hat{y}^{gr_0}(\omega_n)^\dagger\right)^{-1}$$

Proof. By using orthogonal projection and property orthogonal basis, we know that $E\|\Lambda y^{ob} - y^{cy}\|^2$ is minimized if

$$\begin{aligned} 0 &= E[\hat{y}^{ob}(\omega_n)^\dagger(M_n\hat{y}^{ob}(\omega_n) - \hat{y}^{cy}(\omega_n))] \quad \forall n \in I \\ &= E[tr((M_n\hat{y}^{ob}(\omega_n) - \hat{y}^{cy}(\omega_n))\hat{y}^{ob}(\omega_n)^\dagger)] \\ &= tr(M_n\left(\Sigma^{ob}(\omega_n|\mu_0, \theta_0) + \hat{y}^{gr_0}(\omega_n)\hat{y}^{gr_0}(\omega_n)^\dagger\right) - \Sigma^{cy}(\omega_n|\theta_0)) \end{aligned}$$

■

This proposition implies an useful fact. If there exists a linear filter which is designed to extract the business cycle component in minimum mean squared error sense, we can deduce the prior belief on the data generating process that is presupposed by the filter designer. Observe below examples.[6] When we use filter in practice, identical filter is applied to each variables independently. For example, if we use Hodrick Prescott filter with $\lambda = 1,600$, it will be uniformly applied for all variables independently. If we

[5]It is not unbiased

[6]For the more examples of commonly used filter, refer Torben (2009)

describe this filtering by using $\{M_n\}_{n\in I}$ representation, we can write

$$M_n = m(\omega_n)I$$

where $m(\omega_n)$ is a scalar function. Since we knows that minimum mean squared error estimator is represented by

$$M_n = \Sigma^{cy}(\omega_n|\theta_0)\left(\Sigma^{ob}(\omega_n|\mu_0,\theta_0) + \hat{y}^{gr_0}(\omega_n)\hat{y}^{gr_0}(\omega_n)^{\dagger}\right)^{-1}$$

we can think of $m(\omega_n)$ such that

$$m(\omega_n)I = \Sigma^{cy}(\omega_n|\theta_0)\left(\Sigma^{ob}(\omega_n|\mu_0,\theta_0) + \hat{y}^{gr_0}(\omega_n)\hat{y}^{gr_0}(\omega_n)^{\dagger}\right)^{-1}$$

Recall that

$$\Sigma^{ob}(\omega_n|\mu_0,\theta_0) = \Sigma^{gr}(\omega_n|\mu_0) + \Sigma^{cy}(\omega_n|\theta_0)$$

By using this formulation, we can directly observe the tacit assumptions those are imposed by commonly used linear filters.

CASE 1: Hodrick-Prescott filter

This filter suppress the smooth component (growth component in our notation) of observation which is the solution of below minimization problem given λ

$$\min_{\{y_t^{gr}\}} \sum_{t=1}^{T}(y_t^{ob} - y_t^{gr})^2 + \lambda\sum_{t=2}^{T-1}[(y_{t+1}^{gr} - y_t^{gr}) - (y_t^{gr} - y_{t-1}^{gr})]^2$$

In this case, the linear filter Λ is defined by

$$\Lambda y^{ob} = (\delta - h) * y^{ob}$$

where the linear operator $(h*)$ can be represented by $\{H(\omega_n)\}_{n\in I}$ and δ represent the dirac delta.
s.t

$$H(\omega_n) = \frac{1}{4\lambda[1-\cos(\omega_n)]^2 + 1}$$

i.e.

$$m(\omega_n) = \frac{4\lambda[1-\cos(\omega_n)]^2}{4\lambda[1-\cos(\omega_n)]^2 + 1}$$

So, we can conclude that Hodrick Prescott filter tacitly assuming that

$$\Sigma^{cy}(\omega_n|\theta_0) = \frac{4\lambda[1-\cos(\omega_n)]^2}{4\lambda[1-\cos(\omega_n)]^2+1}\left(\Sigma^{cy}(\omega_n|\theta_0) + \Sigma^{gr}(\omega_n|\mu_0) + \hat{y}^{gr_0}(\omega_n)\hat{y}^{gr_0}(\omega_n)^{\dagger}\right)$$

CASE 2: Exponential smoothing filter

This filter is similar to Hodrick Prescott filter. Difference between this filtering and Hodrick Prescott filtering is the minimization criteria.

$$\min_{\{y_t^{gr}\}} \sum_{t=1}^{T}(y_t^{ob} - y_t^{gr})^2 + \lambda \sum_{t=2}^{T}(y_t^{gr} - y_{t-1}^{gr})^2$$

In this case,

$$m(\omega_n) = \frac{2\lambda[1-\cos(\omega_n)]}{2\lambda[1-\cos(\omega_n)]+1}$$

Therefore, we can conclude that this filter is assuming that

$$\Sigma^{cy}(\omega_n|\theta_0) = \frac{2\lambda[1-\cos(\omega_n)]}{2\lambda[1-\cos(\omega_n)]+1}\left(\Sigma^{cy}(\omega_n|\theta_0) + \Sigma^{gr}(\omega_n|\mu_0) + \hat{y}^{gr_0}(\omega_n)\hat{y}^{gr_0}(\omega_n)^{\dagger}\right)$$

For the Hodrick-Prescott filter and exponential smoother, we can see that the variance of growth component is assumed to be bigger in low frequency than high frequency. It is a stochastic version of low frequency or smoothness assumption of growth component. The extreme case of such assumption is imposed by ideal high pass filter.

CASE 3: Ideal high-pass filter

The frequency response of this filter is characteristic function where cut off frequency $f^{\text{cut off}}$ is given. Specifically,

$$m(\omega_n) = \begin{cases} 1 & \text{if} \quad |\omega_n| > f^{\text{cut off}} \\ 0 & \text{o.w} \end{cases}$$

Since $\Sigma^{cy}(\omega_n)(\Sigma^{gr}(\omega_n) + \Sigma^{cy}(\omega_n)) = \boldsymbol{I}\mathbf{1}_{|\omega_n|>f^{\text{cut off}}}(\omega_n)$, this filter tacitly assuming that

$$\begin{aligned} |\Sigma^{cy}(\omega_n|\theta_0)| = 0 \qquad & \text{if} \qquad |\omega_n| \le f^{\text{cut off}} \\ |\Sigma^{gr}(\omega_n|\mu_0) + \hat{y}^{gr_0}(\omega_n)\hat{y}^{gr_0}(\omega_n)^{\dagger}| = 0 \qquad & \text{if} \qquad |\omega_n| > f^{\text{cut off}} \end{aligned}$$

Observe that if above assumption is true, mean squared error of this filter is zero.

CASE 4: Beveridge-Nelson filter

The rational behind the Beveridge-Nelson filter, that was introduced by the seminal work Beveridge and Nelson(1981), is the idea that the growth component is the long run forecastings of observation at a given time. More precisely, we can write this conjecture as follow

$$y_t^{gr} = E\left[\lim_{\tau\to\infty} y_{t+\tau}^{ob}|\Omega_t\right]$$

where Ω_t is the information set which is available at time t. Specifically, Beveridge-Nelson filter assumed ARIMA$(p,1,q)$ on data generating process that is

$$w_{k,t} = \sum_{n=1}^{p} \psi_{k,n} w_{k,t-n} + \sum_{m=1}^{q} \theta_{k,m}\varepsilon_{k,t-m} + \varepsilon_{k,t}$$

where

$$w_{k,t} = y_{k,t}^{ob} - y_{k,t-1}^{ob}$$

and $y_{k,t}$ represents the k the element of a l dimensional real vector.
Defind

$$\begin{aligned}\psi_k(\omega_n) &= 1 - \sum_{m=1}^{p} \psi_{k,m} e^{i\omega_n m}\\ \theta_k(\omega_n) &= 1 + \sum_{m=1}^{q} \theta_{k,m} e^{i\omega_n m}\end{aligned}$$

Put $\Psi(\omega_n)$ and $\Theta(\omega_n)$ be the diagonal matrices with thier k the diagonal elements be $\psi_k(\omega_n)$ and $\theta_k(\omega_n)$ respectively. Then we can derive that Beveridge-Nelson filter is equivalent to define

$$\begin{aligned}M_n &= I - \Theta(0)\Psi(\omega_n)\left(\Theta(\omega_n)\Psi(0)\right)^{-1}\\ &= \left(\Theta(\omega_n)\Psi(0) - \Theta(0)\Psi(\omega_n)\right)\left(\Theta(\omega_n)\Psi(0)\right)^{-1}\end{aligned}$$

i.e.

$$\begin{aligned}&\Sigma^{cy}(\omega_n|\theta_0)\left(\left(\Theta(\omega_n)\Psi(0)\right)\right) =\\ &\quad \left(\Theta(\omega_n)\Psi(0) - \Theta(0)\Psi(\omega_n)\right)\left(\Sigma^{cy}(\omega_n|\theta_0) + \Sigma^{gr}(\omega_n|\mu_0) + \hat{y}^{gr_0}(\omega_n)\hat{y}^{gr_0}(\omega_n)^{\dagger}\right)\end{aligned}$$

Actually, it is not possible to use this filter if we can not pin down the matrices $\Theta(\omega_n)$ and $\Psi(\omega_n)$. So that we can fix those values in matrices reasonable manner, we need to estimate the ARIMA model in advance. It is a sort of iterative procedure if we use this filter to estimate the business cycle component. Based on this observation, it is natural to extend the discussion to the case when filter is sequentially updated. One of such case will be discussed in next section.

3.3 EM algorithm interpretation of filtering

So far, we derived the conclusion that a filter is equivalent to impose prior belief. As an extension, it is natural to consider the case when such belief is sequentially updated. For the analysis in this part, we assume that $\hat{y}^{gr_0}(\omega_n) = 0$ or the case when observation centered at zero by using prior knowledge on deterministic growth term y^{gr_0}. i.e we will consider $\{\hat{y}^{ob_0}(\omega_n)\}_{n\in I}$. In this case, if prior belief on $\Sigma^{cy}(\omega_n|\theta)$ is parameterized by θ, then we can find that maximum likelihood estimation which uses pre-filtered data is equivalent to iterate EM algorithm only one time which start from $\hat{\theta}^{(0)}$. Consider below conjecture on updating scheme.
Write

$$M_n(\mu_0, \hat{\theta}^{(k)}) = \Sigma^{cy}(\omega_n|\hat{\theta}^{(k)})\Sigma^{ob}(\omega_n|\mu_0, \hat{\theta}^{(k)})^{-1} \qquad \text{where} \qquad k = 1, 2, \dots$$

i.e. $M_n(\mu_0, \hat{\theta}^{(k)})$ is the minimum mean squared error filter which is based on the belief that $\theta_0 = \hat{\theta}^{(k)}$. Surely, we cannot sure if maximum likelihood estimation based on the pre-filtered data is authentic if this belief is not true since likelihood function is distorted at the true θ. Note that the quasi maximum likelihood estimation based on pre-filtered data maximize distorted likelihood function. And put this quasi maximum likelihood estimator be $\hat{\theta}^{(1)}$ and find minimum mean squared error filter $M_n(\mu_0, \hat{\theta}^{(1)})$. Iterating this process until the estimation converges. We can conjecture that this process will converges to authentic maximum likelihood estimator i.e $argmax_\theta L(\theta|\{\hat{y}^{ob}(\omega_n)\}_{n\in I})$ under sufficient regularity condition. Actually, this conjecture is valid under the sufficient regularity conditions[7] on $L(\cdot|\cdot)$ are imposed.

Proposition 5. *We can find a sequence of filter* $\{M_n(\mu_0, \hat{\theta}^{(k)})\}_{n\in I}$ *for* $k =$

[7]For the details of regularity condition, refer Jeff Wu(1983)

$1, 2, \ldots$ *given* $\hat{\theta}^{(0)}$ *that is equivalent to EM algorithm if* $L(\cdot|\cdot)$ *satisfies the usual regularity conditions for the convergence of EM algorithm.*

Proof.

$$
\begin{aligned}
& E[\log L(\theta|\mu_0, \hat{y}^{cy}(\omega_n))|\mu_0, \hat{\theta}^{(k)}, \hat{y}^{ob}(\omega_n)] \\
= {} & E\left[-\log \pi^l det\left(\Sigma^{cy}(\omega_n|\theta)\right) - tr\left[\Sigma^{cy}(\omega_n|\theta)^{-1}\hat{y}^{cy}(\omega_n)\hat{y}^{cy}(\omega_n)^{\dagger}\right]|\mu_0, \hat{\theta}^{(k)}, \hat{y}^{ob}(\omega_n)\right] \\
& = -\log \pi^l det(\Sigma^{cy}(\omega_n|\theta) - tr\left[\Sigma^{cy}(\omega_n|\theta)^{-1}M_n(\mu_0, \hat{\theta}^{(k)})^2\hat{y}^{ob}(\omega_n)\hat{y}^{ob}(\omega_n)^{\dagger}\right]
\end{aligned}
$$

For all $k = 0, 1, 2, \ldots,$ and $n \in I^+$. Therefore, we can find the sequence of filter that is equivalent to EM algorithm by updating the minimum mean squared error filter. ■

As you can derive from in above proposition, if $L(\cdot|\cdot)$ satisfies regularity conditions for the EM algorithm convergence, maximum likelihood estimation using above sequence of filter will converge to the result of maximum likelihood estimation based on likelihood function as below.

$$
\begin{aligned}
L(\theta|\{\hat{y}^{ob}(\omega_n)\}_{n\in I}, \mu_0) \quad = \quad & \prod_{n\in I^+} \frac{1}{\pi^l \det(\Sigma^{ob}(\omega_n|\mu_0, \theta)} \\
\times \quad & e^{-tr[(\Sigma^{ob}(\omega_n|\mu_0,\theta))^{-1}\hat{y}^{ob}(\omega_n)\hat{y}^{ob}(\omega_n)^{\dagger}]}
\end{aligned}
$$

To sum up, using pre-filtered data to find maximum likelihood estimation is equivalent to one step iteration of EM algorithm and estimation can be improved by using above likelihood function directly.

4 Numerical example

In this section, we will explore the result of numerical experiment to observe the effect of filtering on estimation. To serve this purpose, plausible data set which contains both of growth component and cyclical fluctuation was generated. Based on the randomly generated data, we will observe the distortion of log likelihood function caused by filter and not using filter. Based on the observation, we will compare the result.

4.1 Data generation

The component of interest, business cycle component is assumed to have some distingtive characteristic: cyclical fluctuation. So that we can generate cyclically fluctuating random process, we assume belows.

$$\Phi_1^{cy}(\theta) = \begin{pmatrix} \theta & 0.1812 & 0.6761 \\ \theta & 0.6761 & -0.1812 \\ -0.7 & \theta & \theta \end{pmatrix} \quad , \quad \Phi_0^{cy}(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

And for the simplicity, put $\varepsilon^{cy} \sim N(0, I)$ and $A_0 = I, A_k = 0$ for all $k \neq 0$. Actually, $\theta = 0$ is assumed to generate the data. i.e. $\theta_0 = 0$. Therefore, we can write this process as belows

$$y_{t+1}^{cy} = \Phi_1^{cy}(\theta_0) y_t^{cy} + \varepsilon_{t+1}^{cy}$$

We can easily deduce that this process will generate the cyclical fluctuation with certain dominante frequency band if we notice that $\Phi_1^{cy}(\theta_0)$ is a rotation matrix multiplied by 0.7. Specifically, $\frac{1}{0.7}\Phi_1^{cy}(\theta_0)$ represent the matrix that makes clockwise, left handed rotation with Euler angles $(\phi, \theta, \psi) = (\frac{\pi}{3}, \frac{\pi}{2}, \frac{\pi}{4})$ with x, y, z convention[8]. 0.7 is multiplied to make this system stable. For example, the impluse response of y^{cy} caused by impulse $(1, 0, 0)$ is as figure 1 which shows the cyclicical fluctuation.

Smoothly fluctuating growth part can be considered as the underlying fluctuation that caused by fundamental changes in economics. For example, if we are observing an closed economy which heavily depends on agricultural production, subtle change in average temperature can cause the smooth fluctuation in output level. Firstly, linear trend was used as a deterministic part. The slope coefficient of linear trend (μ_1, μ_2, μ_3) is set to $(0.009, 0.0095, 0.008)$ and intersection is zero. i.e.

$$y_t^{gro} = \Phi^{gro} \begin{pmatrix} t \\ t \\ t \end{pmatrix}$$

where

$$\Phi^{gro} = \begin{pmatrix} \mu_1 & 0 & 0 \\ 0 & \mu_2 & 0 \\ 0 & 0 & \mu_3 \end{pmatrix}$$

[8]At this moment, θ has nothing to do with our parameter of interest θ. Despite of the risk of confusion, I used θ here because it is a convention.

Also, for the second growth part which is stochastic mean zero part, variance covariance matrix is defined as belows if $\omega_n < \frac{\pi}{20}$ and set to zero otherwise.

$$\begin{pmatrix} \Sigma_X(n|\mu) & \Sigma_{XY}(n|\mu) \\ \Sigma_{XY}(n|\mu) & \Sigma_Y(n|\mu) \end{pmatrix} = \begin{pmatrix} \frac{4\pi}{1000}I & 0 \\ 0 & \frac{4\pi}{1000}I \end{pmatrix}$$

Also, the independence of (X_n, Y_n) in n is assumed so that below holds.

$$\hat{y}^{gro}(\omega_n) \quad \sim \quad CN(0, I) \qquad \text{if} \quad |\omega_n| < \frac{\pi}{20}$$

1,000 observations were generated according to the random data generating process described above. Figure 2 is the plot of the first element of each of generated variable vectors. At a first glance at the time domain representation, it is not easy to point out the difference between each component except for which is more or less spiky. However, as you can see from figure 3, the amplitude of frequency domain representation shows that each of component has its own dominant frequency band. Recall that the idea of Hodrick-Prescott filtering and band-pass filtering is to suppress or amplify certain frequency so that we can accentuate the charactereistics of business cycle component.

4.2 Likelihood function

The purpose of this numerical experiment is to find out what really happens if we use a filtered data especially if the filter is based on the wrong assumptions. As an experiment apparatus, log likelihood function is used. Log likelihood function was calculated based on i) with pre-filtered data, ii) true business cycle data, iii) filtered-free approach. The wrong assumption to make this filtering be pathological is $\theta_0 = 10$. The domain of calculation is $\theta \in [-0.5, 0.5]$ where $\theta_0 = 0$.

Figure 4 is the result of simulation. The first thing you can observe is the fact that frequency domain maximum likelihood estimation will nicely behave if we can use true business cycle component data. Second, if we use the pre-filtered data with the wrong assumption, distortion on likelihood function can be severe and the estimation suffers from error. Finally, we can see that the estimation without filtering can be a better strategy since likelihood function in this case is less distortive and picking point is closer to true value than filtered data case.

To summarize the result of numerical example, the ill-conditioned filter can cause severe distortion on estimation and not using filter can be a better strategy.

5 Conclusion

We have been through the briefly analysis on the estimation of DSGE model and filtering issues so far and it is time to summarize the result of this study. Recall the three objectives of this study what is stated at the introduction of this paper. Firstly, we verified that there exists a correspondence between a filter and a prior belief on data generating process. More precisely, if we consider a filter as a minimum mean squared error estimator, we can deduce the prior belief on data generating process that is presupposed by the filter. By using such correspondence we can codify and clarify the prior belief tacitly implied by a filter. i.e. we can think of a filter as a set of prior belief and using pre-filtered data to estimate the DSGE model is equivalent to impose such arbitrary belief on estimation.

Secondly, we analytically proved that estimating DSGE model by using pre-filtered data can be distortive even if prior belief on growth component is true. It is trivial consequence if we realize the fundamental difference in purpose of filtering and estimation. The purpose of using filter is to distinguish the desirable component from undesirable one. To serve this purpose, we need to assume the prior knowledge on both of undesirable and desirable component. On the other hand, the purpose of estimation is to make an inference on the truth by using available information when we are not sure what the truth exactly is. In other word, if we assume the perfect prior knowledge on truth, it is more reasonable to stick to that belief than estimate the truth. In this context, it is existentially bizarre to estimate the DSGE model by using pre-filtered data.

Thirdly, the estimation method that does not use filter was devised. This method is devoid of the noxious effect caused by ill conditioned prior belief on business cycle component that is tacitly imposed by a filter. Furthermore, under the sufficient regularity conditions, we found out that the result of maximum likelihood estimates without filter is converge to the pre-filtered data based maximum likelihood estimates when we sequentially update the belief what conditions the filter.

Beside the findings of this study, the limitation comes from the setting of this study. We prevent the interaction between business cycle component and growth component to make analysis more tractable. However, this assumption needs to be relaxed in further research. Suppose that there is an investor who is considering buying a stock. If he or she observed that the stock price has risen yesterday, the possible cause is either persistent growth in firms prospective earning or transient cyclical fluctuation of the market. Since investor will take the both possibilities into account when he or she makes a decision, growth component and cyclical component can be not independent. Therefore, it is required to endogenize such correlation structure in further research.

A Figures

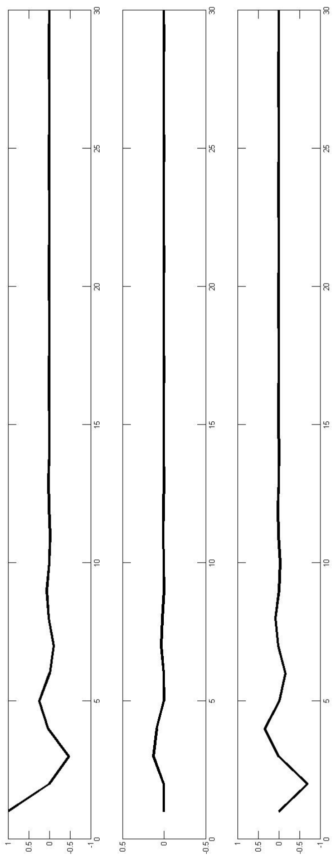

Figure 1: Impulse response of business cycle component caused by shock $(1, 0, 0)$

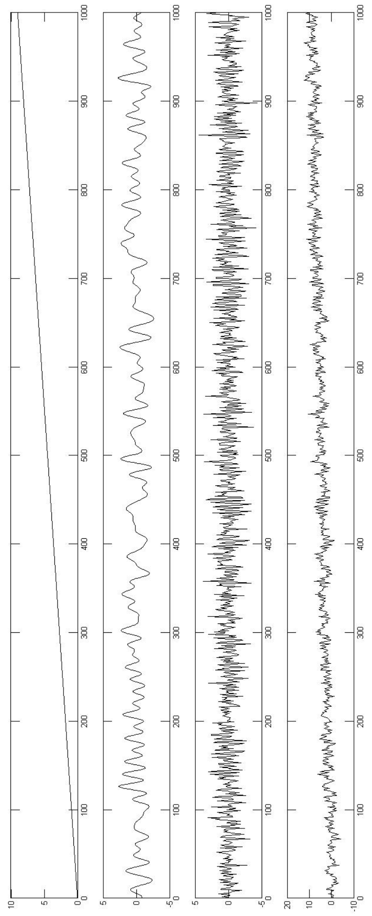

Figure 2: first element of $y^{gr_0}, y^{gr_1}, y^{cy}$ and y^{ob}

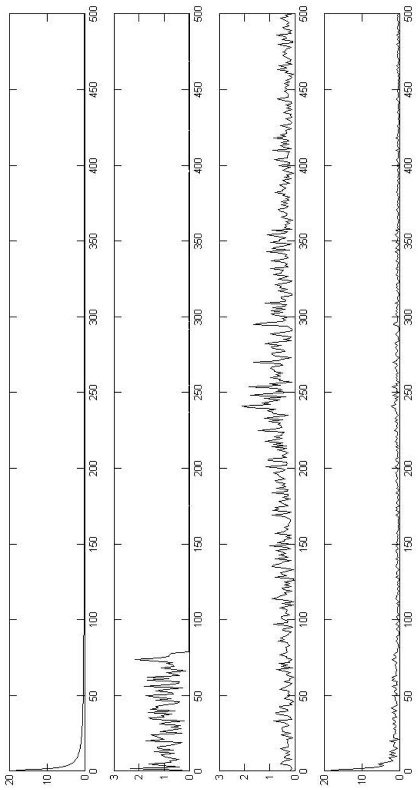

Figure 3: amplitude of frequency domain representation of first element of $y^{gr_0}, y^{gr_1}, y^{cy}$ and y^{ob}

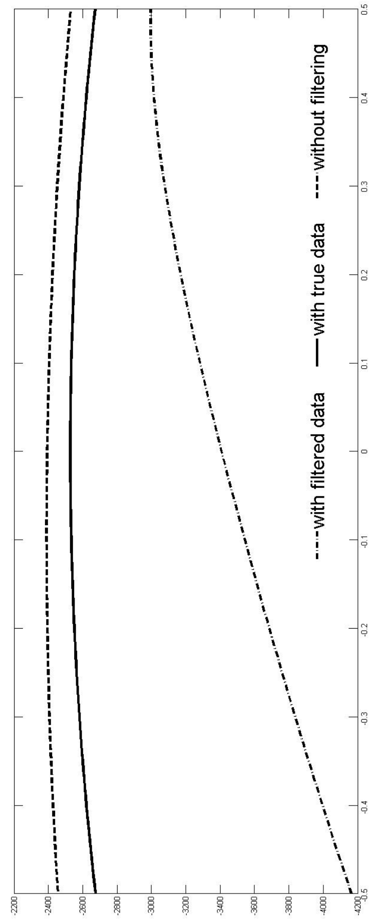

Figure 4: log likelihood function from each case

References

[1] C.F. Jeff Wu (1983). On the convergence properties of the EM algorithm *The Annals of Statistics*, Vol.11, No.1

[2] Christopher A. Sims (2001). Solving linear rational expections models *Computational Economics*, Vol.20, No.1-20

[3] Stephen Beveridge and Charles R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle *Journal of Monetary Economics*, Vol. 7, Issue 2

[4] Edward James Hannan (1970). *Multiple time series* Wiley: New York

[5] Fabio Canova (1998). Detrending and business cycle facts *Journal of Monetary Economics*, Vol.41, Issue 3

[6] Jesús Navarro-Moreno et al.(2009). Estimation of improper complex-valued random signals in colored noise by using the hilbert space theory *IEEE, Transactions on Information theory*, Vol. 55, No.6

[7] Torben Mark Pedersen (2009). The Hodrick-Prescott filter, the Slutzky effect, and the distortionary effect of filters *Journal of Economic Dynamics and Control*, Vol. 25, Issue 8